TERA-OPS PROCESSING FOR ATR

Suraphol Udomkesmalee,[*] Curtis Padgett,[*] David Zhu,[*] Gerald Lung,[*] Ayanna Howard,[*] David Ludwig,[†] and Maj. George Moretti[‡]

ABSTRACT

A three-dimensional microelectronic device (3DANN-R) capable of performing general image convolution at the speed of 10^{12} operations/second (ops) in a volume of less than 1.5 cubic centimeter has been successfully built under the BMDO/JPL VIGILANTE program. 3DANN-R was developed in partnership with Irvine Sensors Corp., Costa Mesa, California. 3DANN-R is a sugar-cube-sized, low-power image convolution engine that in its core computation circuitry is capable of performing 64 image convolutions with large (64x64) windows at video frame rates.

In this paper, we explore potential applications of 3DANN-R such as target recognition, SAR and hyperspectral data processing, and general machine vision using real data and discuss technical challenges for providing deployable systems for BMDO surveillance and interceptor programs.

INTRODUCTION

The Viewing Imager/Gimbaled Instrumentation Laboratory and Analog Neural Three-dimensional processing Experiment (VIGILANTE) program [1]-[2] has successfully developed a three-dimension microelectronic device (3DANN-R) capable of performing general image convolution at the speed of 10^{12} operations/second (ops) in a volume of less than 1.5 cubic centimeter. 3DANN-R was developed in partnership with Irvine Sensors Corp., Costa Mesa, California. 3DANN-R is a sugar-cube-sized, low-power image convolution engine that in its core computation circuitry is capable of performing 64 image convolutions with large (64x64) windows at video frame rates (see Fig. 1). Fast image convolution is fundamental to almost all techniques used in processing images acquired from either passive or active sensors. Numerous operations, including template matching, morphology, classification, and even many model-based matching approaches can be solved using correctly assembled convolution results. By being able to simultaneously generate 64 transformations of an original image, new capability in synthetic image generation, analysis/fusion, and semantic interpretation can be realized with human-like efficiency. 3DANN-R has been proven to function properly and it can already be produced in small quantities for less than $70,000 per unit. The device might cost only a few hundred dollars under mass production.

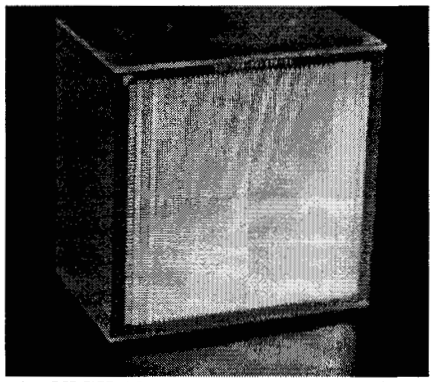


Figure 1: VIGILANTE 3DANN-R 3D Convolution processor, containing 64 row convolver IC's capable of 1 TeraOPS in a 1.4cm x 1.45cm x .75 cm/5W package.

For demonstration purposes, the 3DANN-R requires a PC-case and associated computer circuitry to convert all 64 channels to digital form through a single, very high-speed, analog-digital converter (ADC) and then place those data values in an integrated memory that could be directly interfaced to any computer architecture (see Fig. 2). This VIGILANTE processing architecture creates a real-time, low-mass/power microelectronic visual center capable of transforming raw imagery into a myriad of synthetic images useful for a variety of machine vision and automatic target recognition (ATR) applications.

In this paper, we explore potential applications of 3DANN-R such as target recognition, synthetic aperture radar (SAR) and hyperspectral data processing, and general machine vision using real data and discuss technical challenges for providing deployable systems for BMDO surveillance and interceptor programs.

[*] Jet Propulsion Laboratory, Pasadena, California
[†] Irvine Sensors Corporation, Costa Mesa, California
[‡] Ballistic Missile Organization, Washington, DC

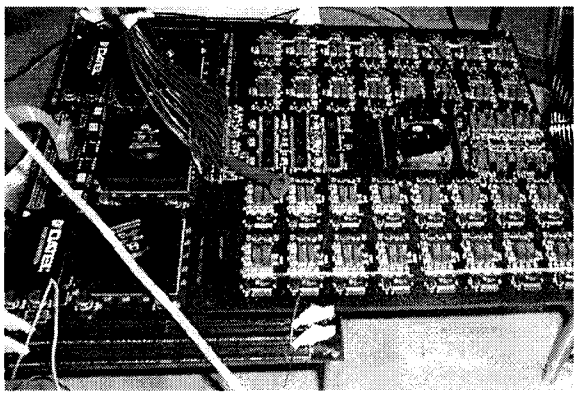


Figure 2: The PC-case support electronics for 3DANN-R consisted of 64 ADCs (10 bits) with low-noise pre-amplifiers, 3 FPGAs (100,000 gates) and 128 Mbytes of memory.

SYSTEM DESCRIPTION

The system is a combination of a P6-based host computer on a PCI backplane and a PCI expansion chassis that contains the 3DANN-R sugarcube processor, custom high-speed PCI interface I/O cards, SHARC board, and a memory buffer (see Fig. 3).

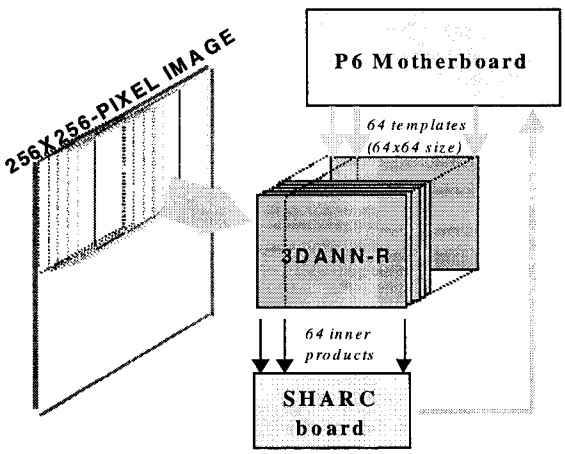
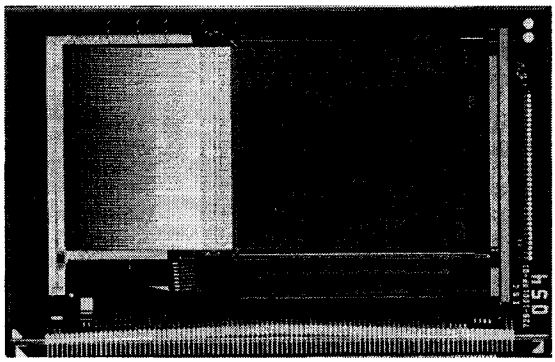
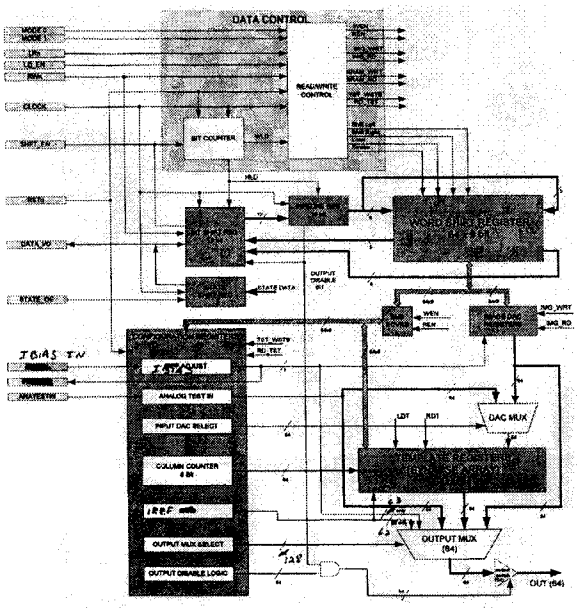


Figure 3: The VIGILANTE processing architecture that orchestrates the data flow from image frame buffer through neural processor also serves as the basis for developing methodologies for ATR applications.

Support electronics for 3DANN-R include a 9"x9" 24-layer printed circuit board comprised of 64 low-noise pre-amplifiers driving 64 10-bit ADCs and a motherboard with three 100,000-gate Field Programmable Gate Arrays (FPGA) and 128 Mbytes of memory (Fig. 2). The input image is stored in a frame buffer that is baselined at 256x256 8-bit pixels frame and can accommodate up to 30 frames per second. The host processor transfers and formats the selected sensor image once every 250 ns. The formatting of data involves rearranging a raster version of the 256x256 image and storing it into 64-byte wide contiguous memory word locations. Every 250ns, a formatted 64-byte row or column image data is loaded into an array of 64x64 D-to-A converters internal to 3DANN-R.



(a)



(b)

Figure 4: The 3DANN-R die chip layout (a) and block diagram (b).

A set of 64 templates (64x64) can be simultaneously handled by 3DANN-R. Each 3DANN-R mixed-signal ASIC (352milsx549mils 0.6μm CMOS), shown in Fig. 4, is a set of 64 row convolvers capable of 10^{12} operations/second (ops) for a stack of 64 [3]. Over 1 million transistors are employed to provide a 9-bit serial digital interface. 4096 8-bit multiplying DACs store templates (filters) that are multiplied with 64 row-templates to produce

one of 64 analog outputs. Outputs of the stack of 64 dies are bussed together to provide current sums from each IC, thus provides an image convolution engine for 2D image processing. A complete set of 64 templates (64x64) can be loaded in 1 ms.

Analog outputs from 3DANN-R every 250 ms are digitized (8-bit resolution) and loaded into a memory buffer for output processing by the SHARC board. Closing the data loop, the host processor can evaluate results from the SHARC board and setup scenarios for ATR.

TARGET RECOGNITION

Since conventional brute-force template matching is usually unreliable in highly cluttered environments (such as in Fig. 5), we have investigated a neural network classification based on eigenvector projections, see Fig. 6. For our experiment using 3DANN-R, we employ directed principal component analysis [4] to generate the generalized eigenvectors for the filter set:

$$\mathbf{S}\,W = \lambda\,\mathbf{R}\,W \qquad (1)$$

where **S** is the covariance matrix for the images with targets, **R** is the covariance matrix for the images without targets, and W is the directed principal components used as the filter set.

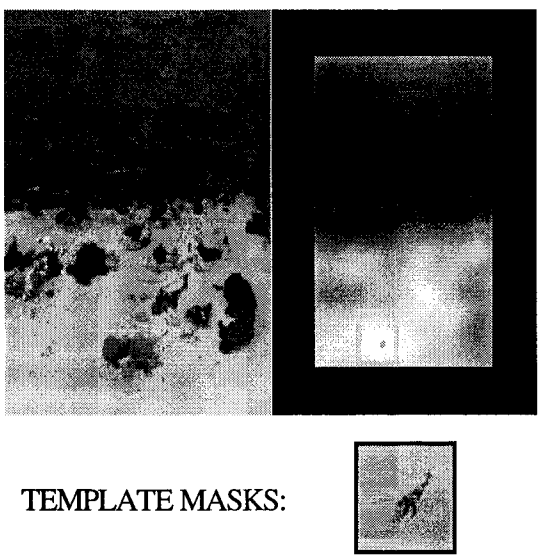


Figure 5: The brute-force template matching with a template mask of the target in a slightly different angle, although provide local maximum corresponding to the target location in the correlation image, generates a false positive result when seeking absolute maximum.

Figure 7 shows the set of the linear filter set used as templates for 3DANN-R processing to produce outputs for each test image frame (see Fig. 8). 16 templates are employed for this particular application.

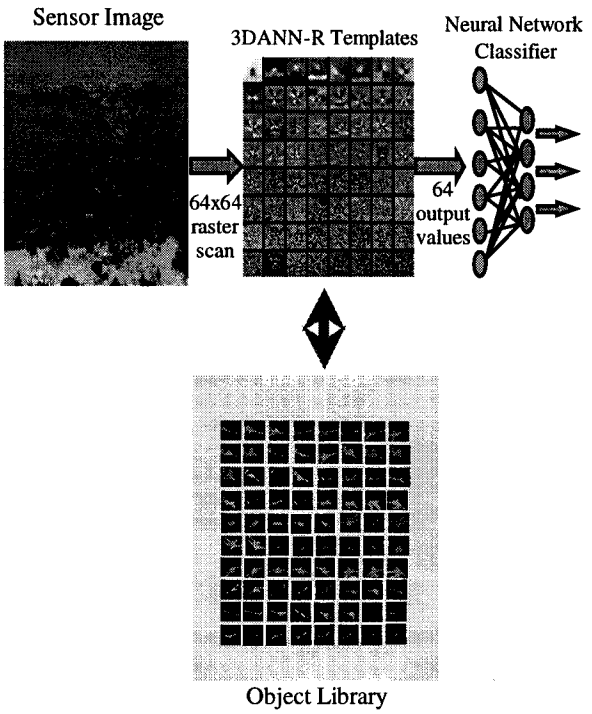


Figure 6: Eigenvector/neural network-based target recognition synthesizes multiple composite filters using Eigenvectors generated from the object library for 3DANN-M processing and classifies corresponding output value with a feed-forward neural network (which can be done in the SHARC board).

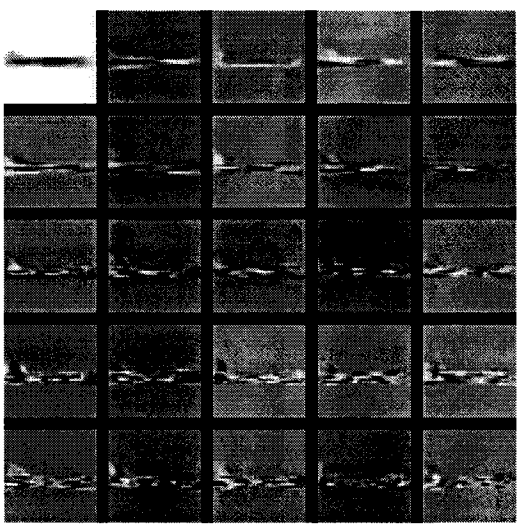


Figure 7: Synthesized filter set (target views and scales vary by 20 degrees and 20%.

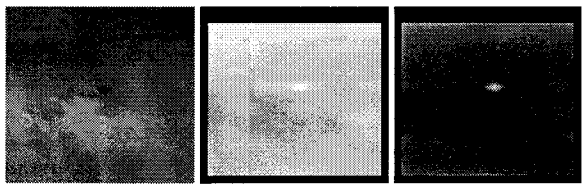


Figure 8: Test image (left) and example outputs (corresponding to the 0^{th} and 13^{th} template of Fig. 7) of cube.

The final steps in our target recognition algorithm involves clustering of each pixel location based on the 16 projected values from the templates (20 clusters) and then employing a specialized *expert* neural network (trained only on examples from its particular cluster) for classification. The networks have been trained off-line using back propogation, and the output from this classifier (shown in Fig. 9) demonstrates recognition of the target in a clutered environment through multiple viewing angles and scales. Detailed performance analysis of this technique can be found in [3].

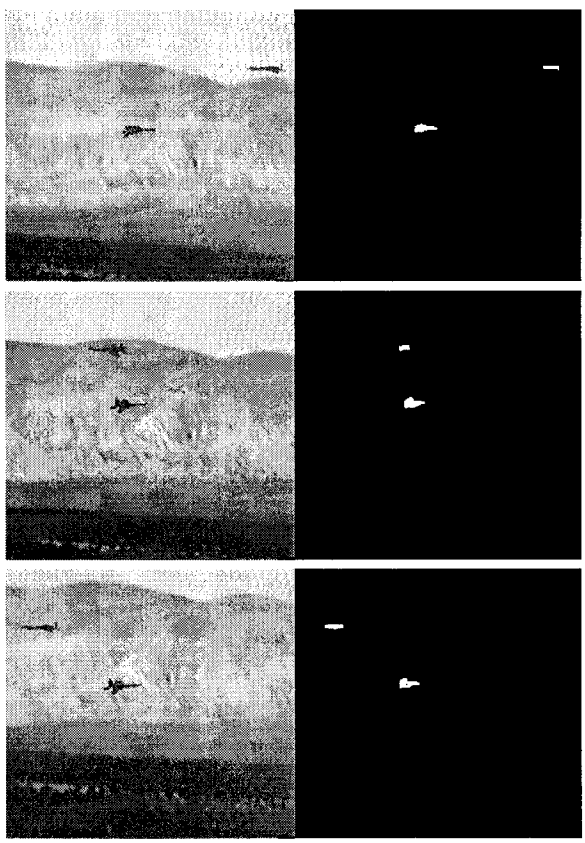


Figure 9: Classifier outputs for a sequence of test images.

SAR DATA PROCESSING

Raw data from SAR are provided in the Fourier domain where a set of complex numbers corresponding to the Fourier-transform of the radar signals for various azimuth angles (α) is given for each range value (d). To derive SAR images from the raw data, a series of 1-dimensional inverse Fourier transform operation is required:

$$f_d(\alpha) = \Sigma_\omega [\mathbf{R}_d(\omega)+j\mathbf{I}_d(\omega)] [\cos(2\pi\omega\alpha/N) + j\sin(2\pi\omega\alpha/N)] \quad (2)$$

where $\mathbf{R}_d(\omega)$ and $\mathbf{I}_d(\omega)$ are the real and imaginary part of the raw data and $f_d(\alpha)$ is the processed SAR data.

The test data shown in Fig. 10 contains 256 α-values for each of the 64 d-values.

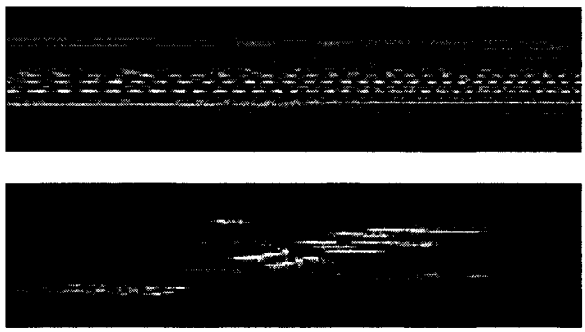


Figure 10: The magnitude of the complex (FFT) SAR data and the inverted distance vs. angle magnitude response of a Boeing727.

Since $f_d(\alpha)$ is a real-valued function, $f_d(\alpha)$ is determined using:

$$f_d(\alpha)^2 = (\Sigma_\omega[\mathbf{R}_d(\omega)\cos(2\pi\omega\alpha/N)] - \Sigma_\omega[\mathbf{I}_d(\omega)\sin(2\pi\omega\alpha/N)])^2 + (\Sigma_\omega[\mathbf{I}_d(\omega)\cos(2\pi\omega\alpha/N)] + \Sigma_\omega[\mathbf{R}_d(\omega)\sin(2\pi\omega\alpha/N)])^2 \quad (3)$$

Using 3DANN-R, Eq. (3) for the above test data must be implemented in 64-value blocks and can be accomplished as follows:

1) Load into channel#1 {$\cos(2\pi\omega\alpha)$, α=0,1,...7, ω=0,1,...,255} (note that each α value will occupy 4 rows), channel#2 {$\cos(2\pi\omega\alpha)$, α=8,9,...,15, ω=0,1,...,255}, ... , channel#32 {$\cos(2\pi\omega\alpha)$, α=248,249,...,255, ω=0,1,...,255}.
2) Load into channel#33-channel#64 the $\sin(2\pi\omega\alpha)$components.
3) Load the 64×64 input image of 3DANN-R with {$\mathbf{R}_d(\omega)$, d=1, ω=0,1,...,255} in the first 4rows and zero for the remaining rows.
4) Readout channels 1-32 for $\mathbf{R}_d\cos(\alpha)$ and 33-64 for $\mathbf{R}_d\sin(\alpha)$, shift down the input image by 4 rows and repeat this step 8 times.

5) Compile $\Sigma_\omega[\mathbf{R}_d(\omega)\cos(2\pi\omega\alpha/N)]$ and $\Sigma_\omega[\mathbf{R}_d(\omega)\sin(2\pi\omega\alpha/N)]$ outputs, and repeat Steps 2-4 for d=2,3,...,64.
6) Repeat Steps 3-5 for $\Sigma_\omega[\mathbf{I}_d(\omega)\cos(2\pi\omega\alpha/N)]$ and $\Sigma_\omega[\mathbf{I}_d(\omega)\cos(2\pi\omega\alpha/N)]$.
7) Generate $f_d(\alpha)$ by taking the square root of the final summation shown in Eq. (3).

Figures 11 and 12 show the result of the calculation performed digitally and the corresponding result from the cube respectively. A total of about 1 msec is required to complete the 64x256 SAR image (assuming the sin and cos templates are pre-loaded).

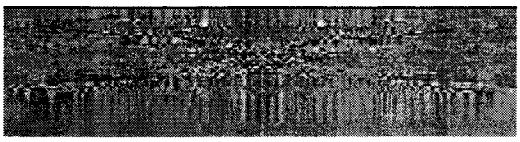


Figure 11: Real component of the output SAR image.

Figure 12: Imaginary component of the output SAR image.

Figure 13: The combined SAR image.

HYPERSPECTRAL DATA PROCESSING

Hyperspectral data provides a spectral signature at each pixel location and creates a data cube structure for a ground map (Fig. 14). In this paper, we use data from the JPL Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) and the algorithm for target spectral recognition described in [5].

AVIRIS is an optical sensor that delivers calibrated images of the upwelling radiance in 224 spectral channels (bands) with wavelength ranging from 400-2500nm using a whiskbroom scan mechanism. The spectral reflectance for each pixel covers a $20m^2$ ground patch area. Data is processed on a per pixel set which represents a series of 1-dimensional array spectral signatures. Recognized target spectra are then labeled for processed pixel locations. The recognition algorithm employs the same generalized eigenvector solution and neural network classifier described in the above target recognition section to derive the filter set for various target classes (prototypes). However, since the data is a 1-dimensional array of 224 elements, procedures for utilization of 3DANN-R described in the SAR data processing section is employed.

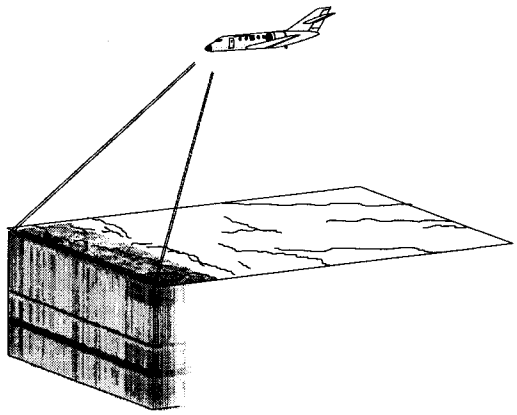


Figure 14: Hyperspectral data structure.

Figures 15-17 shows the result of the classification outputs of the AVIRIS Cuprite copper mine scene.

Generating the 20 template values takes 2 micro-seconds per pixel (8 bit data resolution). To process 12 bit data (used in the simulations of [5]) requires that the data be split in to a low and high order term effectively doubling the per pixel evaluation time.

Figure 15: The AVIRIS Cuprite copper mine scene.

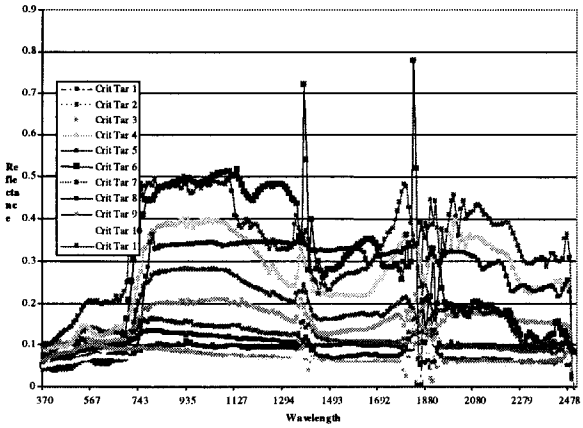


Figure 16: Example spectral prototypes (11 prototypes) used as target spectra.

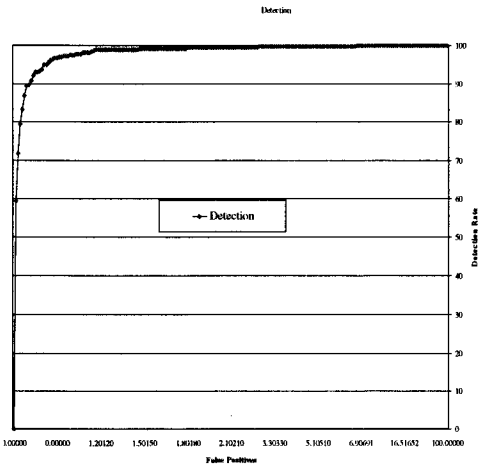


Figure 17: Detection vs. false positive rates for targets mixed in to Cuprite spectra at 10%. The graph shows a false positive rate of less than 1 in 100,000 with a detection rate of over 90%.

MACHINE VISION

The 3DANN-R cube can also be used to perform many of the typical algorithms that support other machine vision applications. Edge images and feature detectors [6], for instance, convolve kernels (usually with a dimension of 7 or 9) with an image and then combine these results with some saliency criteria resulting in a single image highlighting a particular feature.

Figure 18 shows the results from several standard kernels applied to a scene of Venice. The output on a per pixel basis for the edge image was derived from the 4 kernels associated with the Robinson's edge detector using a **max** operation. The other images were the result of a single template passed over the image. The entire set of images was generated in a single pass through the Venice image using only 6 templates (58 are available to do other work). Examples of other kernels or filters processed by the cube include Gaussian, Gabor wavelets, gradients, and mean. These are standard pre-processing steps for numerous vision applications.

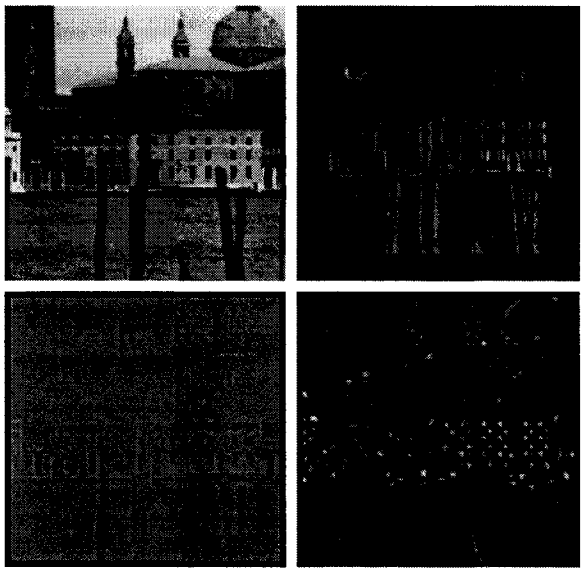


Figure 18: Image operations on scene of Venice (left) using standard kernels implemented on the cube. The top right image is the resultant edge map from Robinson's compass filters. The bottom left image is the output from a laplacian kernel and the bottom right is a simple corner detector.

FUTURE WORK

Continuing work to evaluate system applications of 3DANN-R are needed. Algorithms based on the VIGILANTE architecture to provide sensor fusion for combinations of radar, IR, visible, UV, and hyperspectral sensors must be developed for national missile defense applications. Overall architecture and operational issues dealing with on-line vs. off-line generation of target/background library, logistics of system training, and priority assignments when dealing with multiple targets must also be addressed.

From the system hardware perspective, further optimization is needed for potential field deployments of this tera-ops processor. Although 3DANN-R has impressive capabilities, it currently requires a PC-case of associated computer circuitry to function as the complete VIGILANTE image processing system. In spite of its size, the segregation of the various signal-processing functions onto separate printed circuit boards offered maximum testability of both the 3DANN-R module and the support electronics. Further integration of memory, control and conditioning circuitry into the design

would greatly reduce the size of overall systems, while expanding possible applications by reducing system noise and potentially increasing speed. Currently, to feed image data, upload templates, synchronize output data, and map each output image into the appropriate memory area, the current VIGILANTE system employs several circuit boards. This effort would eliminate these circuit boards and replace them with a small, fast package containing components that push the state of the art in fast mixed mode (analog-digital) processing.

The next generation 3DANN IC will extend the utility of the current design by integrating existing off-chip operations onto the IC. A 0.1-0.2μm CMOS process will be used to realize the added functionality of the new design. The core of the ASIC will remain as a 64X64 8-bit multiplying DAC array driven by a 64X1 9-bit input image DAC. A programmable gain stage (one per column of 64 templates), adjustable through the serial I/O command input will be designed. Together with the off-chip FPGA, this circuit will provide the necessary Automatic Gain Control (AGC) for dynamic in-situ gain adjustments as the templates are changed. An order of magnitude increase in dynamic range can be achieved with an insignificant increase in power and die area. Inclusion of an on-chip AGC offers the added benefit of simplifying the off-chip FPGA and memory control design.

The most significant development will be the addition of a 64-channel transimpedance amplifier followed by a high-speed sample-and-hold and analog multiplexor. Although present on every slice of a 64-stacked-IC module, this analog array of amplifiers will only be enabled on a single die hence effectively summing the output signals from all 64 slices of the module. The on-chip high-speed analog multiplexor and sample-and-hold will reduce the existing 64 output count to only eight hence reducing the required number of external 10-bit ADCs from 64 to only eight thus greatly simplifying the required external circuitry.

Based on the foundation of a working IC design while incorporating the external functions of a proven system architecture, the new 3DANN mixed-mode ASIC will serve as the baseline element for future processors capable of 10-30 teraops and enable a 1000 frames/second multi-sensor ATR system.

CONCLUSIONS

The VIGILANTE vision processing device is economical, extremely small, low power, and ultra-fast which can be used in space, deployed on the ground, or flown on UAV's. It allows autonomous detection, classification, and tracking of targets and items of interest in the midst of enormous data streams. Thus, large-scale networks of intelligence collection systems could assemble the "big picture" using only processed results, reducing the required bandwidth for detailed wide-area surveillance. Furthermore, such system would be useful in centralized ground stations to reduce the analyst workload required to reduce large imagery sets into exploitable information.

In this paper, we have demonstrated the flexibility and practicality of 3DANN-R in various ATR applications. 3DANN-R is at least three orders of magnitude in processing speed better than currently available microprocessors, and assuming Moore's law, it will take at least 15 years for any semiconductor device to catch up.

ACKNOWLEDGMENTS

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the Ballistic Missile Defense Organization through an agreement with the National Aeronautics and Space Administration.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

The authors would also like to thank Lt.Col. Steve Suddarth (AFRL) for his enthusiastic support of the VIGILANTE project and for providing us with SAR data.

REFERENCES

1. S. Udomkesmalee, A. Thakoor, C. Padgett, T. Daud, W.-C. Fang, and S. C. Suddarth, "VIGILANTE: An advanced sensing/processing testbed for ATR applications," *Proc. SPIE*, vol. 3069, pp. 82-93, 1997.
2. S. Udomkesmalee and S. C. Suddarth, "VIGILANTE: Ultrafast Smart Sensor for Target Recognition and Precision Tracking in a Simulated CMD Scenario," *7th Annual AIAA/BMDO Technology Readiness Conf.*, August 1998.
3. J. Carson and D. Ludwig, "TerOps Convolution Engine for Image Processing, *Proc. SPIE*, vol. 4041, April 2000.
4. A. Howard, C. Padgett, and K. Brown, "Real Time Intelligent Target Detection and Analysis with Machine Vision," *Proceedings of 2000 World Automation Congress*, June 2000.

5. A. Howard and C. Padgett, "Intelligent Target Detection in Hyperspectral Imagery," *13th Applied Geologic Remote Sensing Conference*, March 1999.
6. R. Haralick and L. Shapiro. *Computer and Robot Vision, Vol 1*, Addison-Wesley, 1992.